AD_____

Award Number: DAMD17-02-1-0634

TITLE: Spectral Analysis of Breast Cancer on Tissue Microarrys: Seeing Beyond Morphology

PRINCIPAL INVESTIGATOR: David L. Rimm, M.D., Ph.D.

CONTRACTING ORGANIZATION: Yale University School of Medicine
New Haven, CT 06520-8047

REPORT DATE: May 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2003	Annual (15 Apr 2002 - 14 Apr 2003)

4. TITLE AND SUBTITLE
Spectral Analysis of Breast Cancer on Tissue Microarrays: Seeing Beyond Morphology

5. FUNDING NUMBERS
DAMD17-02-1-0634

6. AUTHOR(S)
David L. Rimm, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Yale University School of Medicine
New Haven, CT 06520-8047

E-Mail: David.rimm@yale.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)
No Abstract Provided.

14. SUBJECT TERMS
No subject terms provided.

15. NUMBER OF PAGES
16

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Conventional analysis of breast cancer specimens has largely be based on the microscopic appearance of the tumor. Surprisingly, the microscopic and molecular analysis of tissue have ignored color, a potentially tremendous source of information. Preliminary data suggests that the information content of the spectra of tissue is as high, or higher than that obtained from conventional spatial morphology. Recently, the combination of new optical technologies (spectral imaging) and vastly improved computer power has evolved such that quantitative spectral analysis can be done on each pixel of a complex histologic image. The purpose of this project was to test the hypothesis that spectral analysis will provide diagnostic and prognostic information beyond that attainable from conventional morphology using the same starting material, a stained histology or cytology slide. To test that hypothesis we proposed a three-fold approach, First, we will determine the ability of spectral analysis to distinguish benign from malignant breast tumors. Secondly, we will determine if spectral information can segment patient cohorts based on outcome (in a manner analogous to the way conventional morphology uses histologic and nuclear grade). Finally, we will assess the whether the spectral signatures can be used in a broader fashion to aid diagnosis in cytologic specimens.

Body

The original approved statement of work was as follows:

Tasks/Aims:
Aim 1: To use spectral analysis to classify benign from malignant in breast tissue specimens
Aim 2: To use spectral analysis to attempt to stratify breast cancers with respect to outcome in a manner comparable to histologic/nuclear grade, clinical stage, or prognostic marker results.
Aim 3. To use spectral signatures to classify cytologic breast fine needle aspiration specimens.

Year 1:
1. Construct benign/malignant tissue array from existing tissue collections using approximately 250 cases of breast cancer and associated normal tissue. Assemble associated databases.
2. Hire fellow and train on spectral analysis software.

3. Begin pilot analysis of breast tissue, defining benign/malignant spectral signatures.
4. Select and/or prepare cytologic FNA specimens for Aim 3 cohort.

Year 2:
1. Collect and analyze data on benign/malignant array and define spectral signatures.
2. Completion of preliminary work and publication of first description of methods for spectral-based pathologic analysis.
3. Begin analysis of the classification potential of spectral signatures by collection of spectral data from arrays where the stage, grade, and outcome information is used in selection of machine training regions. This will probably include integration and further training on new software (as it is developed)
4. Optimization/standardization of staining protocols for cytology specimens.
5. Collection of spectral data and data analysis of control cytologic specimens.

Year 3:
1. Completion and publication of first efforts on spectral classification.
2. Continuation of analysis of the classifying capacity of spectral signatures by optimization of the information used in selection of machine training regions. This may include integration of new software as it is developed.
3. Application of spectral signature to cytologic specimens to attempt to stratify on the basis of benign vs malignant, but then also to classify "atypical" cases based on their spectral profile. This may also require further training on cytology-specific modifications of the software.

To date we have essentially completed all tasks targeted for year one and made some progress on other tasks. Specifically, benign and malignant breast cancer tissue microarrays have been constructed and the relevant clinical follow-up information has been collected. Raj Jaganath learned to use the Varispec™ device and software and collected image stacks on a 20 benign and malignant spots. He was trained and assisted in this effort by Dr. Richard Levenson, a key consultant to the project. The images were reviewed and annotated by Tolgay Ocal, a collaborating expert breast pathologist, then sent to Neil Harvey at Los Alamos National Labs for Genie-based software analysis. Genie is Unix-based software produced at Los Alamos, based on the genetic

algorithm concept using both spatial and spectral data as the computational basis. The analysis software is not yet usable by general pathologist. However, Dr. Harvey used the annotations of Dr. Ocal to define normal from malignant and then constructed training sets based on the spectral profiles of a series of spots. Then a second series of spots was selected as a test set. The result was that over 87% of all cancerous nuclei pixels were correctly identified while less than 7% of normal tissue was incorrectly labeled as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labeled less than 1% of pixels as cancer. Thus, although this is preliminary data, it is very promising. It has been presented at the SPIE Biomedical Imaging Conference of 2003 and published in the meeting proceedings. The publication is included in the appendix.

Finally, Raj has prepared a series of 80 breast FNA specimens for analysis by prepping and staining the specimens in identical manner. Some of these FNAs have been annotated by either Dr. Ocal or Dr. Cesar Angeletti and image collection on this FNA data set is about to begin.

Key Research Accomplishments:

1. Completion of Initial Breast Tissue Microarrays for Malignant vs Normal and Outcome-based analysis

2. Completion of Spectral Image stack acquisition for Malignant vs Normal series.

3. Completion of Analysis of Malignant vs Normal series and construction of training and out-of-training set analyses

Reportable outcomes:

An abstract and presentation describing the combination of spectral and spatial analysis showing good classifying ability to distinguish normal breast tissue from malignant tissue (see appendix)

Conclusions:

Preliminary results suggest there is sufficient information attainable from the combination of spectral and spatial data, using genetic algorithms, to classify malignancy in breast cancer. We will now progress to more difficult challenges of distinguishing

benign lesions from malignant lesions and correlation of spectral/spatial features with tumor behavior

References

Appendix

Neal R. Harvey, Richard M. Levenson, David L. Rimm (2003) Investigation of Automated Feature Extraction Techniques for Applications in Cancer Detection from Multispectral Histopathology Images Proc SPIE 2003

Investigation of Automated Feature Extraction Techniques for Applications in Cancer Detection from Multispectral Histopathology Images

Neal R. Harvey*[a], Richard M. Levenson[b], David L. Rimm[c]

[a]NIS-2, Los Alamos National Laboratory, Los Alamos, NM, 87545;

[b]Cambridge Research and Instrumentation Inc., 35-B Cabot Road Woburn, MA 01801;

[c]Dept. of Pathology, Yale University School of Medicine, 310 Cedar St., New Haven, CT 06520

ABSTRACT

Recent developments in imaging technology mean that it is now possible to obtain high-resolution histological image data at multiple wavelengths. This allows pathologists to image specimens over a full spectrum, thereby revealing (often subtle) distinctions between different types of tissue. With this type of data, the spectral content of the specimens, combined with quantitative spatial feature characterization may make it possible not only to identify the presence of an abnormality, but also to classify it accurately. However, such are the quantities and complexities of these data, that without new automated techniques to assist in the data analysis, the information contained in the data will remain inaccessible to those who need it. We investigate the application of a recently developed system for the automated analysis of multi-/hyper-spectral satellite image data to the problem of cancer detection from multispectral histopathology image data. The system provides a means for a human expert to provide training data simply by highlighting regions in an image using a computer mouse. Application of these feature extraction techniques to examples of both training and out-of-training-sample data demonstrate that these, as yet unoptimized, techniques already show promise in the discrimination between benign and malignant cells from a variety of samples.

Keywords: multispectral, histopathology, classification, cancer, machine learning

1. INTRODUCTION

In the field of pathology, accuracy in tissue diagnosis is essential to ensure that patients receive the most appropriate, most cost-effective and least toxic therapies. At present, the state of the art for the determination of a pathological diagnosis relies on manual, morphology based analysis of tissue sections, a method largely unchanged since the nineteenth century. Relying largely upon visual pattern recognition of tissue samples, the entire process is subjective, somewhat irreproducible and inefficient in extracting all the information contained in the specimen, especially as related to prognosis and therapy guidance. Recent advances in optical technologies, coupled with improved computer power, mean that it is now possible to extract information beyond the capabilities of the human visual system. We can extend beyond the limitations of the human eye's acuity and the visible spectrum and obtain high-resolution histological image data at multiple wavelengths. These data have the potential for revealing (often subtle) distinctions between different types of tissue that could be useful in determining objective, reproducible disease-classifying information. The spectral content by itself contains a great deal of information, whose value increases greatly when it is combined with the spatial information available. Unfortunately, such are the quantities and complexities of these data, that without new automated techniques to assist in the data analysis, the useful information contained in the data may remain largely inaccessible. Integration of the spectral and spatial information contained in these images using sophisticated but robust statistical techniques should make it possible to obtain disease classifications that are more accurate, objective and reproducible than is possible with existing manual methods.

Here we describe preliminary experiments in which we investigate the application of a recently developed system for the automated analysis of multi-/hyper-spectral satellite and aerial image data to the problem of cancer detection from multispectral histopathology image data. The system, known as GENIE, was originally developed for the military and intelligence community, to provide a means to develop automated feature extraction tools for multi-

* harve@lanl.gov; phone +1 505 667 9077; fax +1 505 665 4414

and hyper-spectral aerial and satellite imagery. The reason for GENIE's development is that while there exist highly-skilled image analysts who are expert at identifying features of interest from complex image data sets, they are limited in number and, being human, have limited capabilities: they have a limited spectral capability (3 channels) and rate at which they can analyze imagery. So, in order to go beyond these limitations, there is a need to develop systems in which the power of modern computers and machine-learning techniques can be brought to bear. Although human analysts are extremely good at finding features of interest within imagery, they are not so good at describing exactly how they are able to do this, and hence, hand-coding algorithms designed for specific tasks is a difficult and often long and expensive process. Thus, we have developed a system whereby a human expert can teach a computer to create algorithms to perform these functions, via a simple graphical user interface, in which a human provides training data to the computer by simply highlighting examples of the features of interest on a computer screen.

One can make certain comparisons between the military and intelligence community and the medical (specfically, pathology) community. They both have a great deal of complex, high-dimensional data (multi- and hyper-spectral satellite and aerial imagery vs multi-spectral histopathology imagery) that they wish to analyze. They both wish to find features of interest within complex backgrounds (e.g. military targets vs cancer cells) and they both have human experts available who are highly skilled at identifying these features (image analysts vs pathologists), but who have limitations with regard to the complexity and quantity of the data which they can analyze. Bearing these similarities in mind, it is not unreasonable to investigate the application of a software system originally developed to address remote-sensing problems to a set of problems in the medical arena.

2. SPECTRAL IMAGING

Spectral imaging microscopy represents a technological advance over visual or RGB-camera-based analyses, providing images at multiple wavelengths and generating precise optical spectra at every pixel. These rich data sets have applications in surgical pathology, multicolor fluorescence and immunohistochemistry. There now exists a variety of technologies for use in combination with microscopy, including tunable filters, Fourier-transform interferometry, line-scanning prism or gratings-based devices, computed tomography, and others based on polarization effects. Mathematical approaches to these complex data sets may then be used to extract maximum possible information from the resulting data.

In the experiments described here, a VariSpec(tm) liquid crystal tunable filter devices (CRI, Inc.)[14] was used. This device can transmit in a number of wavelength ranges (e.g., 400-720 nm or 850-1800 nm with bandwidths typically in the 7 to 20-nm range, although bandwidths as narrow as 0.1 nm have been achieved).

3. AUTOMATED IMAGE ANALYSIS: OVERVIEW OF THE GENIE SYSTEM

The details of GENIE's algorithmic structure have been described previously in the literature,[1–7] so, in the interests of brevity, we provide only a brief overview of our system.

Our particular interest is the pixel-by-pixel classification of multi-spectral images, not only to locate and identify but also to delineate particular features of interest. For the experiments described here, we are interested in distinguishing cancerous (malignant) cells against the background (which includes normal benign cells). Due to the quantities and complexities of the multispectral data with which we are working, the hand-coding of suitable feature-detection algorithms is impractical. We therefore use a supervised learning approach that can, using only a few hand-classified training images, generate image processing pipelines that are capable of distinguishing features of interest from the background. We remark that our approach here is to consider the two-class problem: although many classification applications require the segmentation of an image into a larger number of distinct classes, for our particular problem, our main interest is the simpler problem of identifying a single class (cancer) against a background of "other" classes. GENIE does possess the capability for performing multiple-class classification,[8] but here we did not make use of that functionality.

GENIE employs a classic evolutionary paradigm: a population is maintained of candidate solutions (*chromosomes*), each composed of interchangeable parts (*genes*), and each assessed and assigned a scalar fitness value, based on how well it performs the desired task. After fitness determination, the evolutionary operators of selection, crossover and mutation are applied to the population and the entire process of fitness evaluation, selection, crossover and mutation is iterated until some stopping condition is satisfied.

3.1. Environment

The environment for each individual in the population consists of *data* planes, each of these planes corresponding to a separate spectral channel in the original multi-spectral image, together with a *weight* plane and a *feature* plane. The weight plane identifies those pixels to be used in training – these are all the pixels for which the analyst has provided a class label. The actual delineation of separate feature/class pixels is given by the feature plane.

3.2. Chromosomes and Genes

Each individual *chromosome* in the population consists of a fixed-length string of *genes*. Each gene in GENIE corresponds to a primitive image processing operation. Therefore the entire chromosome describes an algorithm consisting of a sequence of primitive image processing operations.

Each gene used in GENIE takes one or more distinct image planes as input, and produces one or more image planes as output. Input can be taken from any of the data planes in the training data image cube. Output is written to any of a small number of *scratch planes* — temporary workspaces where an image plane can be stored. Genes can also take input from scratch planes, but only if that scratch plane has been written to by another gene earlier in the chromosome sequence.

Our "gene pool" is composed of a set of primitive image processing operators which we consider useful. These include spectral, spatial, logical and thresholding operators.

3.3. Backends

Final classification requires that the algorithm produce a single (discrete) scalar output plane, which identifies, for every pixel, the class to which it has been assigned. We have found it advantageous to adopt a hybrid approach which applies a conventional supervised classifier to a (sub)set of scratch and data planes to produce the final output plane.

To do this, we first select a subset of the scratch and data planes to be *answer planes*. The conventional supervised classifier "backend" uses the answer planes as input and produces a final output classification plane; in principle, we can use any supervised classification technique as the backend, but for the experiments reported here, we used the Fisher linear discriminant[9] as the backend.

3.4. Fitness Evaluation

The fitness of a candidate solution is given by the degree of agreement between the final classification output plane and the training data. It is based on a simple ratio of the total number of incorrectly classified training pixels over all classes to the total number of training pixels over all classes. If we denote the detection rate (fraction of "true" pixels classified correctly) as R_d and the false alarm rate (fraction of "false" pixels classified incorrectly) as R_f, then the fitness F of a candidate solution is given by

$$F = 500(R_d + (1 - R_f)). \tag{1}$$

Thus, a fitness of 1000 indicates a perfect classification result. This fitness score gives equal weighting to type I (true pixel incorrectly labelled as false) and type II (false pixel incorrectly labelled as true) errors. Note a fitness score of 500 can be trivially achieved with a classifier that identifies all pixels as true (or all pixels as false).

4. EXPERIMENTS: CANCER DETECTION

4.1. Tasks

We set GENIE the task of detecting cancerous nuclei in multispectral breast tissue image data. Thus we have a classification problem with two classes: (1) cancerous nuclei and (2) everything else. Therefore, GENIE was given the task of searching for algorithms that would be able to label each pixel within an image as belonging to one or other of these two classes. While our approach here was to consider the two-class problem, we are aware that other applications might require the segmentation of an image into a larger number of distinct classes. In fact, GENIE is capable of addressing multiple-class problems.[8] However, for this study, we only consider the simpler problem of identifying a single class against a background of "other" classes.

4.2. Multispectral data

The construction of tissue microarrays (TMAs) has been previously described and recently reviewed.[10–13] Briefly, formalin-fixed, paraffin-embedded tissue blocks containing breast cancer were retrieved from the archives of the Yale University Department of Pathology. Areas of invasive carcinoma were identified on corresponding hematoxylin-eosin stained slides and the tissue blocks were cored and transferred to a recipient "master" block using a Tissue Microarrayer (Beecher Instruments, Silver Spring, MD). Each core is 0.6 mm wide, spaced 0.7–0.8 mm apart. After cutting of the recipient block and transfer with an adhesive tape to coated slides for subsequent UV cross-linkage (Instrumedics, Inc, Hackensack, NJ), the slides were dipped in a layer of paraffin in order to prevent oxidation (24). Slides were stained with hematoxylin and eosin, were evaluated for quality of the section and then selected for spectral imaging analysis. For these experiments, examples of both breast cancer and normal tissue were selected.

Images were collected at 10 nm intervals between 420 nm–700 nm using a CRI (Cambridge Research Instruments[15]) VariSpec filter, CRI PanKroma acquisition software, a light microscope, and a QImaging Retiga megapixel digital monochrome camera. The process is semi-automated. The image on the CCD is brought into focus while the tunable filter is tuned to 550 nm (a high-contrast part of the spectrum for H & E samples). An autoexpose function then steps the filter through the spectral range, calculating exposure times wavelength-by-wavelength that will cause the brightest pixels to nearly fill their dynamic range (250 counts for an 8-bit, 256-level sensor). Using these exposure times, a stack of images is automatically collected, with the computer tuning the filter and acquiring an image at every wavelength step, resulting in stacks of 29 images for each sample (tissue microarray dot). To remove optical irregularities in the image train (dust on the CCD window for example) and also some variations in intensity across the liquid crystal filter, the images are flat-fielded by dividing (and normalizing for intensity) each plane of the sample image by the corresponding plane of a white stack obtained from a clear area on the same slide. The image stack, consisting of a series of tif images sequentially numbered, is converted into a single ENVI-format data file with separate header, and transferred to Los Alamos via ftp.

4.3. Training data

In order to provide training data, several images were selected, some containing a mixture of both cancerous and normal tissue and some containing only normal, healthy tissue. Of these images, sub-regions were selected that contained suitably representative samples of pixels from both classes: (1) cancerous nuclei and (2) "everything else". For the cancerous nuclei training samples, regions that had a high density of cancerous nuclei were selected. For the "everything else" training samples, regions were selected that had combinations of the kinds of features that are present in that somewhat-broad class. We were careful to select some regions of normal tissue that contained a high density of normal, healthy nuclei, in order to provide some training data samples that could assist GENIE in evolving an algorithm able to successfully disambiguate cancerous from healthy nuclei.

Fig. 1 shows examples of the original image data and the associated training data (labels) provided by the expert. Fig. 1 (a) shows a true color image of one of the images obtained for breast tissue containing cancer. Fig. 1 (b) shows the training data provided by the expert for the data shown in Fig. 1 (a). Pixels labelled as containing cancer are colored green and pixels labelled as normal are colored red. The training data p(red and green) image has been overlaid onto a gray-scale representation of the true-color image shown in Fig. 1 (a). The region enclosing only those pixels in the image used for training is shown by the bounding box. Fig. 1 (c) shows a true color image of one of the images obtained for breast tissue containing only normal tissue. Fig. 1 (d) shows the training data provided by the expert for the data shown in Fig. 1 (c). As with Fig. 1 (b), pixels labelled as containing cancer are colored green and pixels labelled as normal are colored red (note that there are no green pixels in this image). The training data image has been overlaid onto a gray-scale representation of the true-color image shown in Fig. 1 (c), and the region enclosing only those pixels in the image used for training is shown by the bounding box.

5. RESULTS

Fig. 2 shows the results of applying the classification algorithm found by GENIE during its training, to some data. Fig. 2 shows the classification results of applying the algorithm to the data shown in Fig. 1 (a). The pixels labelled by the algorithm as cancer are colored green and those labelled as normal are colored red. The resulting classification (red and green) image has been overlaid onto a gray-scale image of the original data, just as for the training data shown in Fig. 1 (b). Fig. 2 (b) shows a true-color image of a data set containing cancerous and normal tissue that was not seen during training. Fig. 2 (c) shows a true-color image of a data set containing only normal tissue that

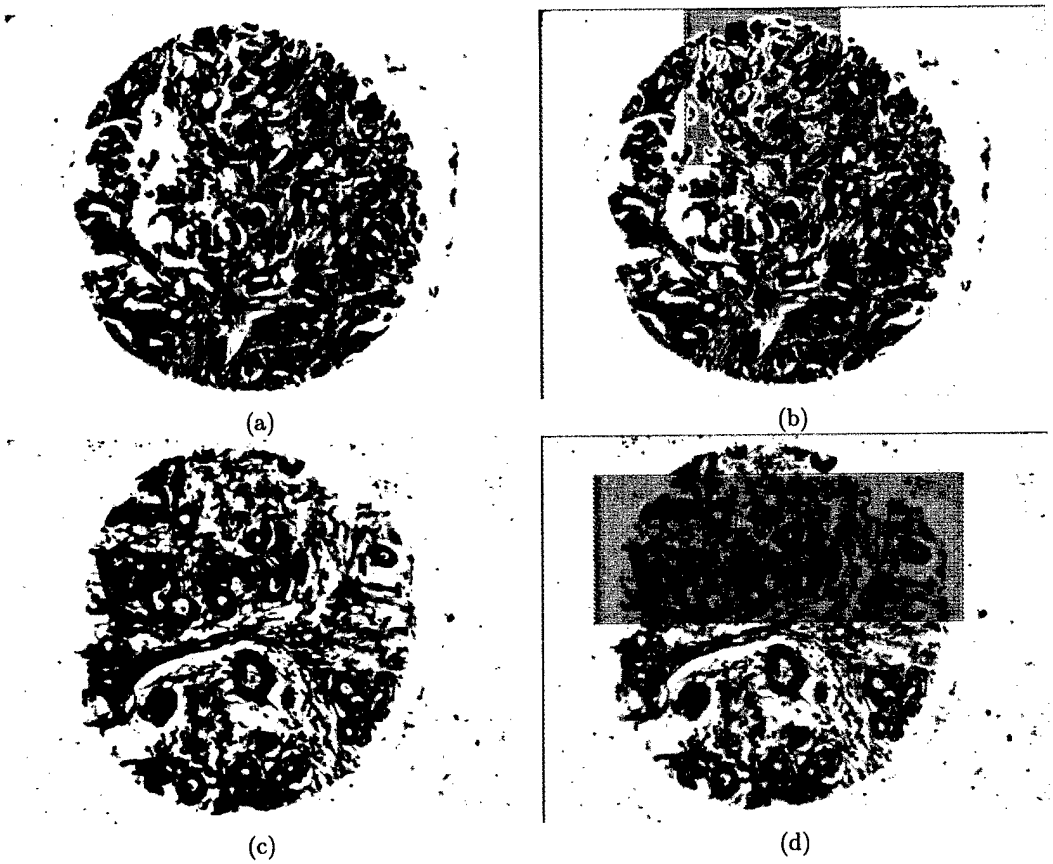

Figure 1. Breast: (a) True color image of one of the data sets obtained from breast tissue containing cancer; (b) Training data provided from this image: Green = Feature (i.e. Cancer), Red = Not Feature (i.e. Non-Cancer); (c) True color image of one of the data sets obtained from breast tissue containing only healthy (non-cancerous) tissue; (d) Training data provided from this image: Notice there are no Green pixels, due to there being no cancer in the image.

was not seen during training. Fig. 2 (d) shows the classification results of applying the algorithm to the data shown in Fig. 1 (c). As before, the pixels labelled by the algorithm as cancer are colored green and those labelled as normal are colored red, and the resulting classification image has been overlaid onto a gray-scale image of the original data.

Table 1 shows the performance of the algorithm found by GENIE during training, as relates to the training data and the entire images, from which the training data was extracted. Column 1 shows the image name. Column 2 shows the number of pixels labelled as cancer that were provided in the training data for each image. Column 3 shows the number of pixels labelled as non-cancer (normal) that were provided in the training data for each image. Column 4 shows the detection rate, DR, (percentage of pixels labelled as cancer in the training data that were labelled correctly as cancer by the algorithm found by GENIE during training) for each image in the training data set. Column 5 shows the false-alarm rate, FAR (percentage of pixels labelled as normal in the training data that were incorrectly labelled as cancer by the algorithm found by GENIE during training) for each image in the training data set. Column 6 shows the total number of pixels labelled as cancer by GENIE's algorithm for the entire image from which the training data was extracted.

Table 2 shows the performance of the algorithm found by GENIE during training, as relates to some testing data - i.e. some image data which was not seen during training (out-of-training-sample data). For these images, in order to be able to assess GENIE's performance in a quantitative manner, an expert provided ground truth for regions in these images, in a similar manner to that provided for the data used in training. Column 1 shows the image name.

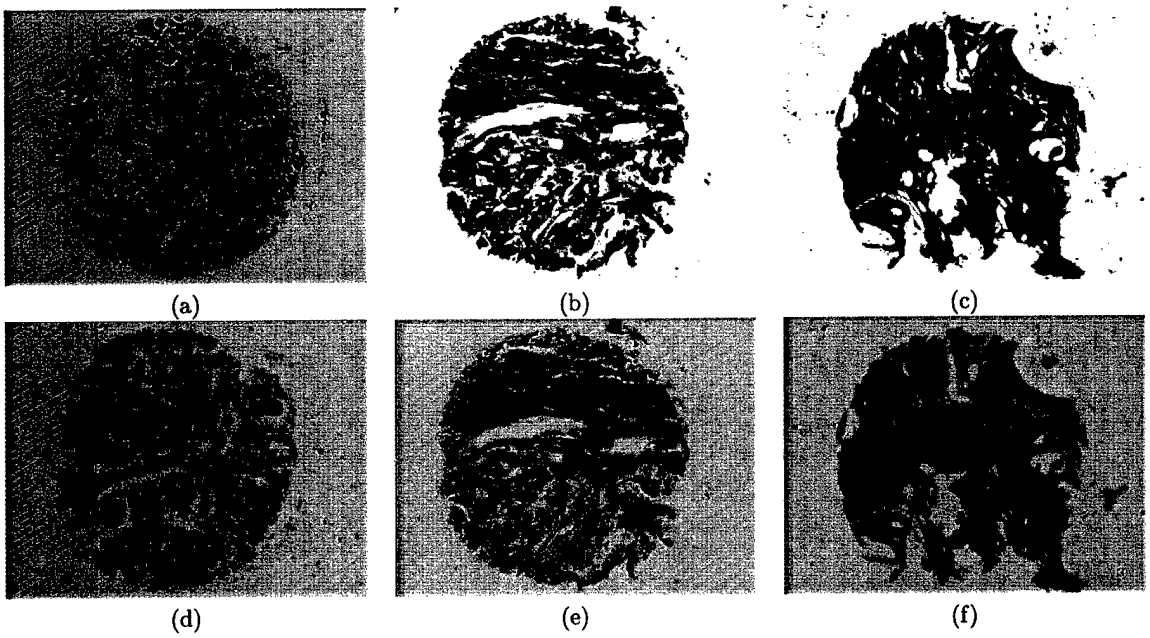

Figure 2. GENIE: Breast (a) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown in Fig. 1 (a); (b) True color image of one of the data sets obtained from breast tissue containing cancer, but not used during training; (c) True color image of one of the data sets obtained from breast tissue containing only healthy (non-cancerous) tissue, but not used during training; (d) Output of GENIE-derived classification algorithm applied to raw multispectral data shown in Fig. 1 (c); (e) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown directly above in (b); (f) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown directly above in (c)

Column 2 shows the number of pixels labelled as cancer by the expert for each image. Column 3 shows the number of pixels labelled as non-cancer (normal) for each image. Column 4 shows the detection rate (DR) for each image in the testing data set. Column 5 shows the false-alarm rate (FAR) for each image in the testing data set. Column 6 shows the total number of pixels labelled as cancer by GENIE's algorithm for the entire image, not just the region labelled by the expert.

Table 2 shows the performance of the algorithm found by GENIE during training, as relates to some testing data - i.e. some image data which was not seen during tyraining (out-of-training-sample data), but for which we don't have expert-provided ground truth. While we don't have expert-provided ground-truth on a pixel-by-pixel basis for these images, we do know, for each image, whether it contains some cancer or whether the image has only normal tissue. Thus, for these images we only provide the total number of pixels labelled as cancer by GENIE's algorithm for the entire image.

6. DISCUSSION

It can be seen, both from the images shown in Fig. 2 and in Tables 1 – 3, that GENIE was able to evolve an algorithm capable of doing a good job of discriminating cancer from non-cancer in the multispectral images used in these experiments. For the training data, for the images that contained both cancerous and non-cancerous (normal) tissue, GENIE was, on average, able to detect over 87% of all cancerous nuclei pixels and only incorrectly labelled less than 7% of normal tissue as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labelled less than 1% of pixels as cancer. For testing data, for which an expert had provided ground-truth, for images that contained a mixture of both cancerous and normal tissue, GENIE, on average, was able to correctly label more

Table 1. Performance of the GENIE-derived classification algorithm found during training applied to training-sample data

Image Name	# Labelled Cancer Pixels (Training)	# Labelled Non-Cancer Pixels (Training)	DR (%)	FAR (%)	Total # Pixels Labelled as Cancer in Result Image
C1-15	25230	147657	70.32	0.17	60742
C2-12	27422	134501	94.72	18.63	138568
C2-9	14871	34187	97.28	0.25	215508
Average	22508	105448	87.44	6.35	138273
N1-8	0	132880	–	0.21	4408
N2-9	0	204768	–	0.21	554
N1-1	0	243120	–	0.36	1305
N4-4	0	335616	–	2.31	18318
Average	0	229096	–	0.77	6146

Table 2. Performance of the GENIE-derived classification algorithm found during training applied to out-of-training-sample data, for which an expert had provided labels, in order to determine out-of-sample performance

Image Name	# Labelled Cancer Pixels (Testing)	# Labelled Non-Cancer Pixels (Testing)	DR (%)	FAR (%)	Total # Pixels Labelled as Cancer in Result
C1-2	12357	83286	48.04	7.85	94274
C2-14	7992	27960	93.23	29.61	300216
C4-9	3880	54773	96.89	18.76	205827
C5-8	4006	73325	90.84	6.99	139858
Average	7059	59836	82.25	15.80	185044
N1-2	0	1.198×10^6	–	0.15	1746
N2-7	0	1.198×10^6	–	0.66	7938
N3-9	0	1.198×10^6	–	0.64	7708
N4-5	0	1.198×10^6	–	0.14	1640
Average	0	1.198×10^6	–	0.40	4758

than 82% of cancerous nuclei pixels and labelled less than 16% of normal tissue incorrectly as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labelled less than 0.5% of pixels as cancer.

It should be noted that the non-nuclei, connective tissue surrounding the cancerous nuclei in the cancer-containing samples is, in fact, not normal tissue. It has its own deviation from normal. It is interesting to note that the algorithm evolved by GENIE labelled this tissue as cancerous. This is hardly surprising. The training data provided from the cancer-containing samples consisted of mostly pixels from cancerous nuclei, with very few samples from the surrounding stroma. However, there were plenty of training samples taken from normal, healthy stroma. Thus, with training samples provided for two classes: malignant nuclei and normal, healthy "everything else", it is understandable that malignant stroma would be significantly different from the training data samples provided for the normal healthy tissue, and would thus be classified into the other "cancerous nuclei" class.

While there was a drop in GENIE's performance, from training data to testing data, for images that contained both cancerous and normal tissue, with the average detection rate going from 87% to 82% and average false-alarm rate going from 7% to 16%, there was actually an improvement in performance, from training data to testing data, for images that contained only normal tissue, with the average false-alarm rate going from 3% to 0.4%.

In general, the algorithm discovered by GENIE does a very good job of discriminating cancer versus normal tissue, both for the data provided in training and for the out-of-training-sample data. There is a large difference (orders of magnitude) between the numbers of pixels classified as being cancer in those images containing cancer

Table 3. Performance of the GENIE-derived classification algorithm found during training applied to out-of-training-sample data, for which no labels had been provided

Image Name	Total # Pixels Labelled as Cancer in Result
C2-5	92385
C2-7	272119
C3-2	169649
C3-4	196517
C5-10	183402
Average	182814
N2-2	292
N2-4	1509
N3-4	686
N3-6	5348
N4-9	4107
Average	2388

compared to those images containing only normal, healthy tissue.

6.1. Further work

GENIE, as it currently stands, despite the promising results shown here, needs much modification in order to be made more generally useful for real applications in pathology. The present suite of operators that make up GENIE's "gene pool" are essentially those which were provided for remote-sensing applications. These operators are not necessarily the most appropriate for the field of pathology. A more targeted group of operators developed from those already developed for such applications in pathology and described in the literature[16,17] would be a good start. In addition, GENIE's current mode of operation, in which the classification is performed on a pixel-by-pixel basis is not ideal. Moving to a higher-level, more object-based classification methodology, would be a better approach. Going even further, beyond providing a simple binary classification indicating the presence or absence of cancer and providing a more detailed classification, such as cancer grade is an additional goal. The other area that needs work is to improve the time taken for training. At present, depending on the amount of training data provided and the complexity of the algorithm space GENIE is set the task of searching, it can take several hours to perform a training run. We aim to be able to reduce this training time to minutes. Our approaches to achieving this goal include parallelisation of the genetic algorithm,[3] implementation of image processing operators in hardware (via FPGAs[18]) and investigation of better, and more efficient search and classification methodologies.[19]

Further work also needs to be undertaken towards a proper validation of the approach, using a far greater volume of data than used in these experiments.

7. CONCLUSIONS

We have shown preliminary investigations into the application of a system originally developed for the automated analysis of satellite image data to the problem of cancer detection from histopathology image data. The results of this work shows great promise, but leaves many questions yet to be answered, and much work to be done.

REFERENCES

1. S.P. Brumby, J. Theiler, S.J. Perkins, N.R. Harvey, J.J. Szymanski, J.J. Bloch and M. Mitchell, "Investigation of Image Feature Extraction by a Genetic Algorithm", in *Proc. SPIE 3812* pp. 24–31 (1999).
2. J. Theiler, N.R. Harvey, S.P. Brumby, J.J. Szymanski, S. Alferink, S. Perkins, R. Porter and J.J. Bloch, "Evolving Retrieval Algorithms with a Genetic Programming Scheme", in *Proc. SPIE 3753*, pp. 416–425 (1999).

3. N.R. Harvey, S.P. Brumby, S.J. Perkins, R.B. Porter, J. Theiler, A.C. Young, J.J. Szymanski, and J.J. Bloch, "Parallel evolution of image processing tools for multispectral imagery", in *Proc. SPIE 4132*, pp.72-82, 2000.
4. S.P. Brumby, N.R. Harvey, S. Perkins, R.B. Porter, J.J. Szymanski, J. Theiler and J.J. Bloch, "A genetic algorithm for combining new and existing image processing tools for multispectral imagery", in *Proc. SPIE 4099*, pp. 480–490, (2000).
5. N.R. Harvey, S. Perkins, S.P. Brumby, J. Theiler, R.B. Porter, A.C. Young, A.K. Varghese, J.J. Szymanski and J. Bloch, "Finding golf courses: The ultra high tech approach", in *Evolutionary Image Analysis, Signal Processing and Telecommunications*, Poli, et al. Springer-Verlag (2000).
6. A.B. Davis, S.P. Brumby, N.R. Harvey, K. Lewis Hirsch, and C.A. Rohde, "Genetic refinement of cloud-masking algorithms for the multi-spectral thermal imager (MTI)" in *Proc. IGARSS 2001*, Sydney, Australia, 9-13 July 2001.
7. N.R. Harvey, J. Theiler, S.P. Brumby, S. Perkins, J.J. Szymanski, J.J. Bloch, R.B. Porter, M. Galassi, and A.C. Young, "Comparison of GENIE and Conventional Supervised Classifiers for Multispectral Image Feature Extraction", in *IEEE Trans. Geoscience and Remote Sensing*, 40:2, (2002), pp. 393–404.
8. N.R. Harvey, J. Theiler, L. Balick, P. Pope, J.J. Szymanski, S.J. Perkins, R.B. Porter, S.P. Brumby, J.J. Bloch, N.A. David, M. Galassi, "Automated Simultaneous Multiple Feature Classification of MTI Data", in *Proc. SPIE 4725*, 99. 346–356 (2002).
9. C.M. Bishop, *Neural Networks for Pattern Recognition*, pp. 105–112, Oxford University Press (1995).
10. D. Rimm, R. Camp, L. Charette, J. Costa, D. Olsen, M. Reiss, "Tissue Microarray: A New Technology for Amplification of Tissue Resources", in *Cancer Journal*, Vol. 7, pp. 24–31, 2001.
11. R.L. Camp, L.A. Charette, D.L. Rimm, "Validation of tissue microarray technology in breast carcinoma", in *Laboratory Investigation*, Vol. 80, Issue 12, pp. 1943–1949, Dec. 2000.
12. G.G. Chung, E.P. Kielhorn, D.L. Rimm. 2002. "Subjective differences in outcome are seen as a function of the immunohistochemical method used on a colorectal cancer tissue microarray" in *Clin. Colorectal Cancer*, Vol. 1, pp. 237–242.
13. G.G. Chung, E. Provost, E.P. Kielhorn, L.A. Charette, B.L. Smith, D.L. Rimm, "Tissue microarray analysis of beta-catenin in colorectal cancer shows nuclear phospho-beta-catenin is associated with a better prognosis", in *Clin. Cancer Res.*, Vol. 7, Issue 12, pp. 4013–4020, Dec. 2001.
14. `http://www.cri-inc.com/instruments/products/imaging_varispec.shtml`
15. `http://www.cri-inc.com/`
16. B. Weyn, G. Van der Wouwer, M. Koprowski, A. Van Daele, K. Dhaene, P. Scheunder, W, Jacob, E. Van Marck, "Value of Morphometry, Texture Analysis, Densitometry and Histometry in the Differential Diagnosis of Malignant Mesolthelioma", in *Journal of Pathology*, Vol. 189, Issue 4, pp. 581–589, Dec. 1999.
17. G.L. Mutter, J.P. Baak, C.P. Crum, R.M. Richart, A. Ferenczy, W.C. Faquin, "Endometrial Precancer Diagnosis by Histopathology, Clonal Analysis and COmputerized Morphometry", in *Journal of Pathology*, Vol. 190, Issue, 4, pp. 462–469, Mar. 2000.
18. R. Porter, K. McCabe and N. Bergmann, "An Applications Approach to Evolvable Hardware", in *Proc. of the First NASA/DoD Workshop on Evolvable Hardware*, Pasadena, California, July 19–21, 1999, pp. 170–174.
19. S. Perkins, N.R. Harvey, S.P. Brumby and K. Lacker, "Support Vector Machines for Broad Area Feature Extraction in Remotely Sensed Images", in *Proc. SPIE 4381*, pp. 286–295, 2001.